

HAB (Harmful Algal Bloom) Detection System

COMP47250 Team Software Project Final Report

Daniel Ilyin
School of Computer Science, UCD
Dublin, Ireland
daniel.ilyin@ucdconnect.ie

Sagar Satish Poojary
School of Computer Science, UCD
Dublin, Ireland
email address @ucdconnect.ie

Dharmik Arvind Vara
School of Computer Science, UCD
 Dublin, Ireland
 email address @ucdconnect.ie

Karthika Garikapati
School of Computer Science, UCD
Dublin, Ireland
email address @ucdconnect.ie

Kruthi Jangam
School of Computer Science, UCD
Dublin, Ireland
email address @ucdconnect.ie

Roshan Palem
School of Computer Science, UCD
Dublin, Ireland
email address @ucdconnect.ie

Abstract—Harmful Algal Blooms (HABs) pose serious risks to marine ecosystems, human health, and coastal economies. Yet most traditional detection methods remain slow, expensive, and limited in how much area they can cover. In this project, we developed a smarter, faster alternative: an AI-powered HAB Detection System that uses satellite imagery and deep learning to predict toxic bloom events more effectively.

At the core of our system is a CNN+LSTM model trained on spatiotemporal data-cubes built from satellite-derived environmental data. The model takes in three key modalities chlorophyll-a concentration, sea surface temperature, and remote sensing reflectance captured across a 10-day window. To make the system accessible to a range of users, we designed a tiered architecture: a lightweight Free Tier using a single modality over 5 days, and more advanced tiers combining richer data for higher accuracy.

Users interact with the system through a simple web interface where they can click on a map to get instant predictions for any location. Our best performing model (Tier 2) achieved 92% accuracy, while the lighter tiers offered reasonable performance with reduced data input and compute needs.

By combining geospatial interactivity, deep learning, and flexible design, this system offers a practical step forward in early HAB detection especially for resource-limited agencies that need timely, reliable insights.

environmental problems, where timely action can help prevent serious harm.

The main challenge we faced was figuring out how to model the way key environmental variables shift over both space and time. To solve that, we used a hybrid model combining Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, trained on spatiotemporal datacubes built from satellite imagery. These datacubes cover 10 days of data across three critical variables: chlorophyll-a levels, sea surface temperature, and remote sensing reflectance.

To make the system flexible and widely usable, we also introduced a tiered prediction structure. Users can choose between a lightweight free version which uses minimal data and more advanced tiers that deliver better accuracy using richer inputs. All of this is wrapped into a simple web interface where users can click on a map and get bloom predictions instantly. The result is a system that combines technical depth with real-world accessibility to stay ahead of harmful algal bloom events.

I. INTRODUCTION

Harmful Algal Blooms (HABs) are showing up more frequently and with greater intensity across coastal waters around the world. These toxic blooms, often caused by nutrient runoff and warming ocean temperatures, can do serious damage to marine life, public health, and local economies. In many cases, they've led to large-scale fish deaths, contaminated drinking water, and forced the shutdown of both recreational beaches and fishing operations. Despite the growing threat, most current methods of detecting HABs still rely on manual sampling and lab analysis which are slow, expensive, and not practical for covering large areas in real time.

Our project set out to build a smarter, more scalable approach using satellite data and deep learning. The aim was to design a system that can predict toxic bloom events from space quickly, accurately, and in a way that's easy for people to use. It fits into a larger effort to apply AI to real-world

[illegible]

II. USER STORIES AND SCENARIOS

A. Marine Biologists and Environmental Agencies

1) *Current Challenge:* Environmental researchers and public health officials require rapid access to comprehensive HAB data for scientific analysis and emergency response planning. Traditional monitoring involving field sampling and laboratory analysis creates unacceptable delays for time-critical environmental assessments.

2) *System Interaction:* Research institutions and environmental agencies access Tier 2 services for comprehensive analysis capabilities. Scientists can upload satellite imagery or specify coordinates for detailed spatiotemporal analysis. The system processes complete 10-day temporal windows across all available modalities, delivering 94% accurate predictions with sub 30-second response times and downloadable datasets for further analysis.

3) *Enabled Capabilities:* This transforms environmental monitoring from weeks-long traditional workflows to minutes of analysis. Researchers gain access to 100km² spatial coverage with complete temporal context, enabling rapid scientific assessment and evidence-based recommendations to decision-makers. Data export capabilities support integration with existing research workflows and regulatory reporting requirements.

4) *Technical Innovation:* This scenario highlights our system's ability to convert state-of-the-art spatiotemporal model outputs into accessible, scientifically rigorous datasets while maintaining research-grade accuracy standards. The tiered architecture ensures computational resources scale appropriately with scientific requirements.

B. Commercial Fishing Operations

1) *Current Challenge:* Commercial fishermen face significant economic risks from HAB events that can contaminate catches, damage equipment, and force costly operational delays. Traditional monitoring provides insufficient spatial and temporal coverage for operational planning, often delivering information too late for effective business decisions.

2) *System Interaction:* A commercial fishing operation utilises Tier 1 services for operational planning. Through the dashboard interface, operators input their fishing coordinates and access 7-day predictions with 80% accuracy. The system processes multi-modal satellite data (chlorophyll-a, SST, PAR) through our CNN-LSTM architecture, providing day-by-day toxicity assessments for specific fishing areas.

3) *Enabled Capabilities:* This capability transforms reactive HAB response into proactive business planning. Operators can schedule fishing activities around predicted safe periods, protecting both equipment investments and catch quality. The extended forecast horizon enables strategic deployment of vessels and crew, reducing economic losses from unexpected bloom events.

4) *Technical Innovation:* This use case showcases our system's ability to deliver operational-grade forecasts with the reliability and temporal resolution required for commercial decision-making, addressing the scalability limitations of traditional monitoring approaches.

C. Recreational Water Users

1) *Current Challenge:* Members of the public planning recreational activities near water bodies have no accessible method to assess HAB risks in advance. Traditional monitoring provides sparse coverage with results available only after laboratory analysis, often days after sampling.

2) *System Interaction:* A family planning a weekend lake visit accesses our Free Tier service through the web interface. Using the interactive map, they select their intended location and visit date. The system processes chlorophyll-a concentration data from recent MODIS imagery, applies our CNN-based classification model, and delivers a 5-day toxicity forecast with 70% accuracy within minutes.

3) *Enabled Capabilities:* This interaction transforms previously inaccessible satellite data analysis into an intuitive point-and-click interface. Users receive actionable safety information without requiring technical expertise, enabling proactive decision-making about recreational activities. The color-coded risk visualization (green/yellow/red) provides immediate comprehension of complex environmental data.

4) *Technical Innovation:* This scenario demonstrates our core contribution of democratizing access to advanced spatiotemporal modeling. The system converts raw multi-modal satellite data into clear, location-specific forecasts, solving the fundamental challenge of making environmental risk assessment usable for non-technical stakeholders.

III. STATE OF THE ART AND RELATED WORK

A. Traditional HAB Detection Methods

Traditional HAB monitoring relies on manual water sampling and laboratory analysis, giving highly accurate results but are very costly, geographically limited, and days-long delays make real-time response impossible.

B. Remote Sensing Approaches

Early satellite based detection used chlorophyll-a thresholds, having a broad coverage but having a 30-40% false positive rates due to interferences and natural variability. Spectral shape algorithms improved specificity by exploiting second-derivative features, while anomaly detection using background subtraction reached 75-80% accuracy but struggled with seasonal patterns.

C. Classical Machine Learning

Support Vector Machines trained on MODIS spectral features improved classification over threshold methods but required extensive manual feature engineering. Random Forest classifiers achieved 82-85% accuracy across regions but lacked temporal modelling capabilities. Early spatiotemporal approaches combined ensemble methods with temporal features but were limited to small datasets (<100 events).

D. Deep Learning Architectures

E. HABNet: Current State-of-the-Art

Hill et al. (2020) introduced spatiotemporal datacubes integrating multiple MODIS modalities (chlorophyll-a, SST, PAR, Rrs bands) over 100km×10-day windows. Their hybrid CNN-LSTM architecture achieved 91% detection accuracy and 86% prediction accuracy up to 8 days ahead, significant improvement over traditional methods (60-70%). However, HABNet suffers critical deployment limitations: no user interface, requires deep technical expertise, lacks multi-user architecture, and provides no flexibility for varying computational constraints.

F. Operational Systems

G. User Interface Design

H. Research Contribution

IV. IMPLEMENTATION

A. System Architecture Overview

V. EVALUATION AND RESULTS

A. Dataset Description

B. Experimental Setup

C. Detection Performance

D. Prediction Performance

E. System Performance Evaluation

F. User Evaluation

G. Technical Evaluation

VI. CONCLUSIONS AND FUTURE WORK

REFERENCES

- [1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," *Phil. Trans. Roy. Soc. London*, vol. A247, pp. 529–551, April 1955.
- [2] J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.